

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup>:</b> <b>C12N 15/62, 15/51, 15/63, 9/22, C12P 21/02</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 97/12043</b> <b>(43) International Publication Date:</b> 3 April 1997 (03.04.97)
<b>(21) International Application Number:</b> PCT/US96/14688 <b>(22) International Filing Date:</b> 12 September 1996 (12.09.96) <b>(30) Priority Data:</b> 08/529,169 15 September 1995 (15.09.95) US <b>(71) Applicant:</b> CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US). <b>(72) Inventors:</b> HANG, Jang; 5700 Fern Street, El Cerrito, CA 94530 (US). CHOE, Joonho; 502-1001 Expo Att., Yu Song-gu, Taejon 305-390 (KR). <b>(74) Agents:</b> HARBIN, Alisa, A. et al.; Chiron Corporation, Intellectual Property - R440, P.O. Box 8097, Emeryville, CA 94662-8097 (US).		<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> HCV NS3 PROTEIN FRAGMENTS HAVING HELICASE ACTIVITY AND IMPROVED SOLUBILITY  <b>(57) Abstract</b>  The Hepatitis C Virus (HCV) NS3 protein contains amino acid motifs of a serine proteinase, a nucleotide triphosphatase (NTPase), and an RNA helicase. A carboxy fragment of the HCV NS3 protein was purified and possessed RNA helicase activity. Deletions from the amino terminus resulted in the protein becoming soluble. Deletions from the carboxy terminus do not result in a loss of helicase activity until at least 50 amino acids are deleted. the helicase activity requires ATP and divalent cations such as Mg <sup>2+</sup> and Mn <sup>2+</sup> . The helicase activity was blocked by monoclonal antibody specific to the HCV NS3 protein.		

Thus, a need exists for an effective method for treating HCV infection: currently, there is none.

HCV is a positive strand RNA virus. Upon infection, its genomic RNA produces a large polyprotein that is processed by viral and cellular proteins into at least 10 different viral proteins. Like other positive strand RNA viruses, replication of the positive strand involves initial synthesis of a negative strand RNA. This negative strand RNA, which is a replication intermediate, serves as a template for the production of progeny genomic RNA. This process is believed to be carried out by two or more viral encoded enzymes, including RNA-dependent RNA polymerase and RNA helicase. RNA polymerase copies template RNA for the production of progeny RNA. This enzyme does not synthesize RNA molecules from DNA template.

The RNA helicase unwinds the secondary structure present in the single-strand RNA molecule. The helicase also unwinds the duplex RNA into single-strand forms. Genomic HCV RNA molecules contain extensive secondary structure. Replication intermediates of HCV RNA are believed to be present as duplex RNA consisting of positive and negative strand RNA molecules. The activity of RNA helicase is believed to be crucial to RNA dependent RNA polymerase which requires unwound single stranded RNA molecules as a template. Therefore, the biological activity of helicase is believed to be required for HCV replication.

NS3 proteins of the three genera of the *Flaviviridae* family: flavivirus, pestivirus and HCV, have been shown to have conserved sequence motifs of a serine- type proteinase and of a nucleoside triphosphatase (NTPase)/RNA helicase. One third of the N'-terminal of the HCV NS3 protein has been shown to be a trypsin like serine proteinase which cleaves the NS3-NS4A, NS4A-NS4B, NS4B-NS5A, and NS5A-NS5B junctions. Faila et al., *J. Virol.* 68:3753-3760 (1994). Two thirds of the NS3 C'-terminal fragment has been shown to encode NTPase/RNA helicase activity. Choo et al., *PNAS*, 88:2451-2455 (1991) and Gorbalenya et al., *Nucleic Acids Res.*, 17:4713-4729 (1989). Suzich et al.

HCV NS3 PROTEIN FRAGMENTS HAVING  
HELICASE ACTIVITY AND IMPROVED SOLUBILITY

5     Technical Field

          This invention relates to the molecular biology and virology of the hepatitis C virus (HCV). More specifically, this invention relates to (1) carboxy terminus helicase fragments of the HCV NS3 protein having improved solubility in extraction and assay buffers, (2) methods of expressing the novel NS3 protein helicase fragments having  
10    improved solubility, (3) recombinant NS3 protein helicase fragments having improved solubility; (4) NS3 protein helicase mutant fragments; and (5) methods of using the HCV NS3 protein helicase fragments for screening helicase inhibitors as potential therapeutic agents.

15    Background of the Invention

          Non-A, Non-B hepatitis (NANBH) is a transmissible disease (or family of diseases) that is believed to be virally induced, and is distinguishable from other forms of virus-associated liver disease, such as those caused by hepatitis A virus (HAV), hepatitis B virus (HBV), delta hepatitis virus (HDV), cytomegalovirus (CMV) or Epstein-Barr virus  
20    (EBV). Epidemiologic evidence suggests that there may be three types of NANBH: the water-borne epidemic type; the blood or needle associated type; and the sporadically occurring (community acquired) type. However, the number of causative agents is unknown. Recently, however, a new viral species, hepatitis C virus (HCV) has been identified as the primary (if not only) cause of blood-associated NANBH (BB-NANBH).  
25    See for example, PCT WO89/046699 and WO92/02642; European Patent Specification 318,216-B, and European Patent Application Publication Nos 388,232-A and 398,748-A, each incorporated herein by reference. Hepatitis C appears to be the major form of transfusion-associated hepatitis in a number of countries, including the United States and Japan. There is also evidence implicating HCV in induction of hepatocellular carcinoma.

Figure 3 shows the conserved sequence motif of DEXH box RNA helicase proteins and comparative alignment of the RNA helicase domain of the HCV NS3 protein. The numbers between boxes indicate the distance in amino acids residues.

5                      Figure 4 shows the structure of double strand RNA substrate for RNA helicase assay. The thick line indicates the  $^{32}\text{P}$ -labeled RNA strand. The thin line indicates the unlabeled RNA strand.

                        Figure 5 shows the expression and purification of HCV NS3 from *E. Coli*.  
10      M: protein size markers, Lane 1: Total protein from uninduced cells, Lane 2: Total protein from 3 hr IPTG induced cells, Lane 3: HCV NS3:His-tag fusion protein purified by nickel binding chromatography.

                        Figure 6 shows the results of an RNA helicase assay of the HCV NS3  
15      protein fragments. Lane (-) enzyme: ds RNA without NS3 protein. Lane boiled: ds RNA denatured by heat. Lane 1; Fraction from negative control cell (pET vector only), Lane 2: 3 mM  $\text{Mn}^{2+}$ , Lane 3: no  $\text{Mn}^{2+}$ , Lane 4: 3 mM  $\text{Mg}^{2+}$ , Lane 5: no  $\text{Mg}_2^{+}$ , Lane 6: 3 mM  $\text{KCl}$ , Lane 7: no ATP, Lane 8: 1 mM ATP, Lane 9: preincubation of the NS3 protein with NS3-specific monoclonal antibody, Lanes 10, 11: preincubation of the NS3 protein with  
20      anticonnexin monoclonal antibody at 0.5  $\mu\text{g}$ , 1.0  $\mu\text{g}$  per 20  $\mu\text{l}$ , respectively. Monoclonal antibodies were preincubated with the S3 protein at room temperature for 5 min.

                        Figure 7 shows the activity profiles of the HCV NS3 RNA fragment having  
                        helicase activity with different ATP and divalent cations concentrations. The effects of  
25      cations were tested at two different ATP concentrations (1 mM and 5 mM).

                        Figure 8 shows the activity of various truncated fragments of the HCV NS3 protein.

showed that two thirds of the carboxy terminal fragment of HCV NS3 expressed in *E. coli* had polynucleotide-stimulated NTPase activity. J. Virol., 67:6152-6158 (1993). Gwack et al., in "NTPase Activity of Hepatitis C Virus NS3 Protein Expressed in Insect Cells" Mol. Cells. 5(2): 171-175 (1995), showed two HCV NS3 proteins, p70 and p43, were expressed  
5 in a baculovirus expression system. The p70 showed a specific NTPase activity that was inhibited by NS3 monoclonal antibodies. Warrener et al., "Pestivirus NS3 (p80) Protein Possesses RNA Helicase Activity," J. Virol. 69:1720-1726 (1995), demonstrated that bovine viral diarrhea virus (BVDV) NS3 protein expressed in a baculovirus expression system had a RNA helicase activity. JP 0631 9583A describes the preparation of a helicase  
10 protein encoded by HCV by introducing a HCV helicase gene into the non-essential region of a baculovirus. The helicase amino acid sequence is reported as 1200 through 1500 of the HCV polyprotein. All documents mentioned above are incorporated herein in their entirety by reference.

#### 15 Disclosure of the Invention

We have now invented recombinant HCV NS3 protein fragments having helicase activity and improved solubility, fusion HCV NS3 protein fragments having helicase activity and improved solubility, truncated and altered HCV NS3 protein fragments having helicase activity and improved solubility, and cloning and expression  
20 vectors therefore, and methods for using these protein fragments in screening assays to assess whether a compound is capable of inhibiting RNA helicase activity and thus inhibiting HCV replication.

#### Brief Description of the Drawings

25 Figure 1 shows the sequence of the of the NS3 protein of HCV-1, which is approximately from amino acid 1027 to 1657 of the HCV-1 polyprotein. (SEQ ID NO: 1)

Figure 2 is a schematic presentation of the HCV NS3 protein. The numbers indicate the amino acid positions of the HCV-1 polyprotein.

Viruses: The Togaviridae And Flaviviridae" (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press, 1986), pp. 327-374. Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter of about 40-50 nm. Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections measuring about 5-10 nm in length with terminal knobs about 2 nm in diameter. Typical examples of the family include Yellow Fever virus, West Nile virus, and Dengue Fever virus. They possess positive-stranded RNA genomes (about 11,000 nucleotides) that are slightly larger than that of HCV and encode a polyprotein precursor of about 3500 amino acids. Individual viral proteins are cleaved from this precursor polypeptide.

The genome of HCV appears to be single-stranded RNA containing about 10,000 nucleotides. The genome is positive-stranded, and possesses a continuous translational open reading frame (ORF) that encodes a polyprotein of about 3,000 amino acids. In the ORF, the structural proteins appear to be encoded in approximately the first quarter of the N-terminal region, with the majority of the polyprotein attributed to non-structural proteins. When compared with all known viral sequences, small but significant co-linear homologies are observed with the non-structural proteins of the Flavivirus family, and with the pestiviruses (which are now also considered to be part of the Flavivirus family).

The HCV polyprotein is processed by the host and viral proteases during or after translation. The genetic map of HCV is as follows: from the amino terminus to the carboxy terminus, the nucleocapsid protein (C), the envelope proteins (E1) and (E2), and the non-structural proteins 2, 3, 4 (a+b), and 5 (a+b) (NS2, NS3, NS4, and NS5). Based upon the putative amino acids encoded in the nucleotide sequence of HCV1, a small domain at the extreme N-terminus of the HCV polyprotein appears similar both in size and high content of basic residues to the nucleocapsid protein (C) found at the N-terminus of flaviviral polyproteins. The non-structural proteins 2,3,4, and 5 (NS2-5) of HCV and of yellow fever virus (YFV) appear to have counterparts of similar size and hydropathicity,

Figure 9 shows the various truncated HCV NS3 fragments and whether the fragments showed helicase/NTPase activity.

#### Modes of Carrying Out The Invention

5

##### A. Definitions

The terms "Hepatitis C Virus" and "HCV" refer to the viral species that is the major etiological agent of BB-NANBH, the prototype isolate of which is identified in PCT WO89/046699; EPO publications 318,216, 388,232 and 398,748, and PCT  
10 WO92/02642. "HCV" as used herein includes the pathogenic strains capable of causing hepatitis C, and attenuated strains or defective interfering particles derived therefrom. The HCV genome is comprised of RNA. It is known that RNA-containing viruses have relatively high rates of spontaneous mutation, reportedly on the order of  $10^{-3}$  to  $10^{-4}$  per incorporated nucleotide (Fields & Knipe, "Fundamental Virology" (1986, Raven Press,  
15 N.Y.)). As heterogeneity and fluidity of genotype are inherent characteristics of RNA viruses, there will be multiple strains/isolates, which may be virulent or avirulent, within the HCV species.

Information on several different strains/isolates of HCV is disclosed herein,  
20 particularly strain or isolate CDC/HCVI (also called HCV1). Information from one strain or isolate, such as a partial genomic sequence, is sufficient to allow those skilled in the art using standard techniques to isolate new strains/isolates and to identify whether such new strains/isolates are HCV. Typically, different strains, which may be obtained from a number of human sera (and from different geographical areas), are isolated utilizing the  
25 information from the genomic sequence of HCV1.

HCV is now classified as a new genus of the *Flaviviridae* family of which the other two genera are pestivirus and flavivirus. The Flavivirus family contains a large number of viruses which are small, enveloped pathogens of man. The morphology and  
30 composition of Flavivirus particles are known, and are discussed in M.A. Brinton, in "The

1193 Val Asp Phe Ile Pro Val Glu Asn Leu Glu  
 Thr Thr Met Arg Ser Pro Val Phe Thr Asp  
 Asn Ser Ser Pro Pro Val Val Pro Gln Ser  
 Phe Gln Val Ala His Leu His Ala Pro Thr  
 5 Gly Ser Gly Lys Ser Thr Lys Val Pro Ala  
 Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu  
 Val Leu Asn Pro Ser Val Ala Ala Thr Leu  
 Gly Phe Gly Ala Tyr Met Ser Lys Ala His

10 (Leu)  
 Gly Ile Asp Pro Asn Ile Arg Thr Gly Val  
 Arg Thr Ile Thr Thr Gly Ser Pro Ile Thr  
 Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp  
 Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile  
 15 Ile Ile Cys Asp Glu Cys His Ser Thr Asp  
 Ala Thr Ser Ile Leu Gly Ile Gly Thr Val  
 Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg  
 Leu Val Val Leu Ala Thr Ala Thr Pro Pro  
 Gly Ser Val Thr Val Pro His Pro Asn Ile  
 20 Glu Glu Val Ala Leu Ser Thr Thr Gly Glu  
 Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu  
 Glu Val Ile Lys Gly Gly Arg His Leu Ile  
 Phe Cys His Ser Lys Lys Lys Cys Asp Glu  
 Leu Ala Ala Lys Leu Val Ala Leu Gly Ile  
 25 Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp  
 Val Ser Val Ile Pro Thr Ser Gly Asp Val  
 Val Val Val Ala Thr Asp Ala Leu Met Thr  
 Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile

30 (Tyr)  
 Asp Cys Asn Thr Cys Val Thr Gln Thr Val

(Ser)  
 Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile  
 35 Glu Thr Ile Thr Leu Pro Gln Asp Ala Val  
 Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly  
 Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val  
 Ala Pro Gly Glu Arg Pro Ser Gly Met Phe  
 Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp  
 40 Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro  
 Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr  
 Met Asn Thr Pro Gly Leu Pro Val Cys Gln  
 Asp His Leu Glu Phe Trp Glu Gly Val Phe  
 Thr Gly Leu Thr His Ile Asp Ala His Phe  
 45 Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn  
 Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr



although the amino acid sequences diverge. The region of HCV, which would correspond to the regions of YFV polyprotein that contains the M, E, and NS1 protein, not only differs in sequence, but also appears to be quite different in size and hydropathicity. Thus, while certain domains of the HCV genome may be referred to herein as, for example, E1, E2, or NS2, it should be understood that these designations are for convenience of reference only; there may be considerable differences between the HCV family and flaviviruses that have yet to be appreciated and as these differences surface, domain designations may change.

Due to the evolutionary relationship of the strains or isolates of HCV, putative HCV strains and isolates are identifiable by their homology at the polypeptide level. With respect to the isolates disclosed herein, new HCV strains or isolates are expected to be at least about 40% homologous, some more than about 70% homologous, and some even more than about 80% homologous: some may be more than about 90% homologous at the polypeptide level. The techniques for determining amino acid sequence homology are known in the art. For example, the amino acid sequence may be determined directly and compared to the sequences provided herein. Alternatively the nucleotide sequence of the genomic material of the putative HCV may be determined (usually via a cDNA intermediate), the amino acid sequence encoded therein can be determined, and the corresponding regions compared.

20

The term "NS3 protein fragment showing helicase activity" or "NS3 protein helicase fragment" refers to an enzyme derived from an HCV NS3 protein which exhibits helicase activity, specifically the portion of polypeptide that is encoded in the carboxy two-third terminus of the NS3 domain of the HCV genome. Generally, the portion of the HCV NS3 protein showing protease activity, i.e., that which is found in the amino one-third terminus, has been removed. At least one strain of HCV contains a NS3 protein fragment showing helicase activity believed to be substantially encoded by or within the following sequence of amino acids residues within the NS3 protein fragment i.e.; approximately amino acids 1193 to 1657 of the NS3 protein shown in Figure 1. The sequence of such helicase fragment is depicted below:

30

purification and assay buffers. The NS3 protein generally is insoluble in buffers. When approximately 20 amino acids of helicase N terminus are deleted, the fragments become soluble in buffer. When approximately thirty-five amino acids are deleted, however, the fragments lose both NTPase and helicase activity. It is known that a portion of the NS3 protein at the amino terminus i.e., that beginning around amino acid 1027 exhibits protease activity. Protease activity, however, is not required of the HCV helicases of the invention and, in fact, the amino terminus fragments of NS3 exhibiting protease activity have been deleted from the helicase or fragments of the present invention.

"HCV NS3 fragment helicase analogs" refer to polypeptides which vary from the NS3 carboxy fragment having helicase activity, shown above, by deletion, alteration and/or addition to the amino acid sequence of the native helicase fragment. HCV NS3 helicase fragment analogs include the truncated helicase fragments described above, as well as HCV NS3 fragment helicase mutants and fusion helicase fragments comprising HCV NS3 protein helicase fragments, truncated NS3 protein helicase fragments, or NS3 fragment helicase mutants. Alterations to form HCV NS3 fragment helicase mutants are preferably conservative amino acid substitutions, in which an amino acid is replaced with another naturally-occurring amino acid of similar character. For example, the following substitutions are considered "conservative":

Gly « Ala; Asp « Glu; Val « Ile « Leu;  
Lys « Arg; Asn « Gln; and Phe « Trp « Tyr.

Nonconservative changes are generally substitutions of one of the above amino acids with an amino acid from a different group (e.g., substituting Asn for Glu), or substituting Cys, Met, His, or Pro for any of the above amino acids. Substitutions involving common amino acids are conveniently performed by site specific mutagenesis of an expression vector encoding the desired protein, and subsequent expression of the altered form. One may also alter amino acids by synthetic or semi-synthetic methods. For example, one may convert cysteine or serine residues to selenocysteine by appropriate chemical treatment of the isolated protein. Alternatively, one may incorporate uncommon amino acids in standard *in vitro* protein synthetic methods. Typically, the total number of residues changed, deleted or

5 Val Cys Ala Arg Ala Gln Ala Pro Pro Pro  
Ser Trp Asp Gln Met Trp Lys Cys Leu Ile  
Arg Leu Lys Pro Thr Leu His Gly Pro Thr  
Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln  
Asn Glu Ile Thr Leu Thr His Pro Val Thr  
Lys Tyr Ile Met Thr Cys Met Ser Ala Asp  
Leu Glu Val Val Thr 1658  
(SEQ ID NO: 2)

10 The above N and C termini of the helicase fragment are putative, the actual  
termini being defined by expressing and processing in an appropriate host of a DNA  
construct encoding the entire NS3 domain. It is understood that this sequence may vary  
from strain to strain, as RNA viruses, like HCV, are known to exhibit a great deal of  
variation. Further, the actual N and C termini may vary, as the NS3 protein fragment  
15 showing helicase activity is cleaved from a precursor polypeptide: variations in the helicase  
amino acid sequence can result in different termini for helicase activity. Thus, the amino-  
and carboxy-termini may differ from strain to strain of HCV. A minimum sequence  
necessary for activity does exist and has been determined herein. The sequence of the NS3  
fragment may be truncated at either end or at both by treating an appropriate expression  
20 vector with exonuclease after cleavage with a restriction endonuclease at the 5' or 3' end (or  
both) of the coding sequence to remove any desired number of base pairs. The resulting  
coding polynucleotide is then expressed and the sequence determined. In this manner the  
activity of the resulting product may be correlated with the amino acid sequence: a limited  
series of such experiments (removing progressively greater numbers of base pairs)  
25 determines the minimum internal sequence necessary for helicase activity. The sequence of  
the HCV NS3 fragment may be substantially truncated, particularly at the carboxy terminus  
up to approximately 50 amino acids, with full retention of helicase activity. Successive  
carboxy truncations do eventually result in the loss of helicase activity. Further carboxy  
truncation, at around 135 amino acids results in the loss of NTPase activity. The amino  
30 terminus of the NS3 fragment, i.e., that beginning around 1190 of the HCV-1 amino acid  
sequence may also be truncated to a degree without a loss of helicase activity.  
Surprisingly, an amino terminus truncation to around twenty amino acids of the putative  
helicase domain does, however, result in an increase in the solubility of the fragment in

### B. General Method

The practice of the present invention generally employs conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature.

- 5 See for example J. Sambrook et al, "Molecular Cloning; A Laboratory Manual (1989); "DNA Cloning", Vol. I and II (D.N Glover ed. 1985); "Oligonucleotide Synthesis" (M.J. Gait ed, 1984); "Nucleic Acid Hybridization" (B.D. Hames & S.J. Higgins eds. 1984); "Transcription And Translation" (B.D. Hames & S.J. Higgins eds. 1984); "Animal Cell Culture" (R.I. Freshney ed. 1986); "Immobilized Cells And Enzymes" (IRL Press, 1986);
- 10 B. Perbal, "A Practical Guide To Molecular Cloning" (1984); the series, "Methods In Enzymology" (Academic Press, Inc.); "Gene Transfer Vectors For Mammalian Cells" (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Meth Enzymol (1987) 154 and 155 (Wu and Grossman, and Wu, eds., respectively); Mayer & Walker, eds. (1987), "Immunochemical Methods In Cell And Molecular Biology" (Academic Press,
- 15 London); Scopes, "Protein Purification: Principles And Practice", 2nd Ed (Springer-Verlag, N.Y., 1987); and "Handbook Of Experimental Immunology", volumes I-IV (Weir and Blackwell, eds, 1986).

- Both prokaryotic and eukaryotic host cells are useful for expressing desired
- 20 coding sequences when appropriate control sequences compatible with the designated host are used. Among prokaryotic hosts, *E. coli* is most frequently used. Expression control sequences for prokaryotes include promoters, optionally containing operator portions, and ribosome binding sites. Transfer vectors compatible with prokaryotic hosts are commonly derived from, for example, pBR322, a plasmid containing operons conferring ampicillin
- 25 and tetracycline resistance, and the various pUC vectors, which also contain sequences conferring antibiotic resistance markers. These plasmids are commercially available. The markers may be used to obtain successful transformants by selection. Commonly used prokaryotic control sequences include the T7 bacteriophage promoter (Dunn and Studier, J. Mol. Biol. (1983) 166:477) the b-lactamase (penicillinase) and lactose promoter systems
- 30 (Chang et al, Nature (1977) 198:1056), the tryptophan (*trp*) promoter system (Goeddel et

added to the native sequence in the mutants will be no more than about 20, preferably no more than about 10, and most preferably no more than about 5.

The term fusion protein generally refers to a polypeptide comprising an amino acid sequence drawn from two or more individual proteins. In the present invention, "fusion protein" is used to denote a polypeptide comprising the HCV NS3 helicase fragment, truncate, mutant or a functional portion thereof, fused to a non-HCV protein or polypeptide ("fusion partner"). Fusion proteins are most conveniently produced by expressing of a fused gene, which encodes a portion of one polypeptide at the 5' end and a portion of a different polypeptide at the 3' end, where the different portions are joined in one reading frame which may be expressed in a suitable host. It is presently preferred (although not required) to position the HCV NS3 helicase fragment or analog at the carboxy terminus of the fusion protein, and to employ a functional enzyme fragment at the amino terminus. The HCV NS3 helicase fragment is normally expressed within a large polypeptide. The helicase fragment is not expected to include cell transport signals (e.g., export or secretion signals). Suitable functional enzyme fragments are those polypeptides which exhibit a quantifiable activity when expressed fused to the HCV NS3 helicase fragment. Exemplary enzymes include, without limitation, b-galactosidase (b-gal), b-lactamase, horseradish peroxidase (HRP), glucose oxidase (GO), human superoxide dismutase (hSOD), urease, and the like. These enzymes are convenient because the amount of fusion protein produced can be quantified by means of simple colorimetric assays. Alternatively, one may employ fragments or antigenic proteins, to permit simple detection by metal-binding columns and quantification of fusion proteins using antibodies specific for the fusion partner. The presently preferred fusion partner is six histidine residues at the carboxy terminus.

amino acid polypeptide which is automatically cleaved from the fused protein upon expression. The ubiquitin amino acid sequence is as follows:

5                   Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu  
                   Glu Val Glu Ser Ser Asp Thr Ile Asp Asn Val Lys Ser Lys  
                   Ile Gln Asp Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu  
                   Ile Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg Thr Leu Ser  
                   Asp Tyr Asn Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu  
                   Arg Leu Arg Gly Gly  
                   (SEQ ID NO: 3)

10

See also Ozkaynak et al, Nature (1984) 312:663-66. Polynucleotides encoding the ubiquitin polypeptide may be synthesized by standard methods, for example following the technique of Barr et al, J Biol Chem (1988) 268:1671-78 using an Applied Biosystem 380A DNA synthesizer. Using appropriate linkers, the ubiquitin gene may be  
 15                   inserted into a suitable vector and ligated to a sequence encoding the HCV helicase or a fragment thereof.

In addition, the transcriptional regulatory region and the transcriptional initiation region which are operably linked may be such that they are not naturally  
 20                   associated in the wild-type organism. These systems are described in detail in EPO 120,551, published October 3, 1984; EPO 116,201, published August 22, 1984; and EPO 164,556, published December 18, 1985, all of which are commonly owned with the present invention, and are hereby incorporated herein by reference in full.

25                   Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells, and a number of other cell lines. Suitable promoters for mam-  
 30                   malian cells are also known in the art and include viral promoters such as that from Simian Virus 40 (SV40) (Fiers et al, Nature (1978) 273:113), Rous sarcoma virus (RSV), adeno-  
                   virus (ADV), and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly-A addition sequences. Enhancer sequences which increase expression may also be included, and sequences which promote amplification of the gene

al, Nuc Acids Res (1980) 8:4057) and the lambda-derived P<sub>L</sub> promoter and N gene ribosome binding site (Shimatake et al, Nature (1981) 292:128) and the hybrid lac promoter (De Boer et al, Proc Nat Acad Sci USA (1983) 292:128) derived from sequences of the trp and lac UV5 promoters. The foregoing systems are particularly compatible with  
5 *E. coli*; if desired, other prokaryotic hosts such as strains of *Bacillus* or *Pseudomonas* may be used, with corresponding control sequences.

Eukaryotic hosts include, without limitation, yeast and mammalian cells in culture systems. Yeast expression hosts include *Saccharomyces*, *Klebsiella*, *Pichia*, and  
10 the like. *Saccharomyces cerevisiae* and *Saccharomyces carlsbergensis* and *K. lactis* are the most commonly used yeast hosts, and are convenient fungal hosts. Yeast-compatible vectors carry markers which permit selection of successful transformants by conferring prototrophy to auxotrophic mutants or resistance to heavy metals on wild-type strains. Yeast compatible vectors may employ the 2m origin of replication (Broach et al, Meth  
15 Enzymol (1983) 101:307), the combination of CEN3 and ARS1 or other means for assuring replication, such as sequences which will result in incorporation of an appropriate fragment into the host cell genome. Control sequences for yeast vectors are known in the art and include promoters for the synthesis of glycolytic enzymes (Hess et al, J Adv  
Enzyme Reg (1968) 7:149; Holland et al, Biochem (1978), 17:4900), including the  
20 promoter for 3-phosphoglycerate kinase (R. Hitzeman et al, J Biol Chem (1980) 255:2073). Terminators may also be included, such as those derived from the enolase gene (Holland, J Biol Chem (1981) 256:1385). Particularly useful control systems are those which comprise the glyceraldehyde-3 phosphate dehydrogenase (GAPDH) promoter or alcohol dehydrogenase (ADH) regulatable promoter, terminators also derived from GAPDH, and if secretion  
25 is desired, a leader sequence derived from yeast  $\alpha$ -factor (see U.S. Pat. No. 4,870,008, incorporated herein by reference).

A presently preferred expression system employs the ubiquitin leader as the fusion partner. Copending application USSN 7/390,599 filed 7 August 1989 disclosed vec-  
30 tors for high expression of yeast ubiquitin fusion proteins. Yeast ubiquitin provides a 76

use the strong viral polyhedrin gene promoter to drive expression of heterologous genes. Currently the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373 (see PCT WO89/046699 and USSN 7/456,637). Many other vectors known to those of skill in the art have also been designed for improved expression. These  
5 include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and introduces a BamHI cloning site 32 bp downstream from the ATT; See Luckow and Summers, Virology (1989) 17:31). AcNPV transfer vectors for high level expression of nonfused foreign proteins are described in copending applications PCT WO89/046699 and USSN 7/456,637. A unique BamHI site is located following position -8 with respect to the  
10 translation initiation codon ATG of the polyhedrin gene. There are no cleavage sites for SmaI, PstI, BglII, XbaI or SstI. Good expression of nonfused foreign proteins usually requires foreign genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. The plasmid also contains the polyhedrin polyadenylation signal and the ampicillin-resistance (amp) gene and origin of  
15 replication for selection and propagation in *E. coli*.

Methods for the introduction of heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summer and Smith, Texas Agricultural Experiment Station Bulletin No. 1555; Smith et al, Mol Cell Biol (1983) 3:2156-2165; and  
20 Luckow and Summers, Virology (1989) 17:31). For example, the heterologous DNA can be inserted into a gene such as the polyhedrin gene by homologous recombination, or into a restriction enzyme site engineered into the desired baculovirus gene. The inserted sequences may be those that encode all or varying segments of the polyprotein, or other orfs that encode viral polypeptides. For example, the insert could encode the following  
25 numbers of amino acid segments from the polyprotein: amino acids 1-1078; amino acids 332-662; amino acids 406-662; amino acids 156-328, and amino acids 199-328.

The signals for post-translational modifications, such as signal peptide cleavage, proteolytic cleavage, and phosphorylation, appear to be recognized by insect  
30 cells. The signals required for secretion and nuclear accumulation also appear to be



may also be desirable (for example methotrexate resistance genes). These sequences are known in the art.

5 Vectors suitable for replication in mammalian cells are known in the art, and may include viral replicons, or sequences which insure integration of the appropriate sequences encoding HCV epitopes into the host genome. For example, another vector used to express foreign DNA is Vaccinia virus. In this case the heterologous DNA is inserted into the Vaccinia genome. Techniques for the insertion of foreign DNA into the vaccinia virus genome are known in the art, and may utilize, for example, homologous recombination. The heterologous DNA is generally inserted into a gene which is non-essential to the virus, for example, the thymidine kinase gene (*tk*), which also provides a selectable marker. Plasmid vectors that greatly facilitate the construction of recombinant viruses have been described (see, for example, Mackett et al, J Virol (1984) 49:857; Chakrabarti et al, Mol Cell Biol (1985) 5:3403; Moss, in GENE TRANSFER VECTORS FOR MAMMALIAN  
10 CELLS (Miller and Calos, eds., Cold Spring Harbor Laboratory, NY, 1987), p. 10). Expression of the HCV polypeptide then occurs in cells or animals which are infected with the live recombinant vaccinia virus.

In order to detect whether or not the HCV polypeptide is expressed from the  
20 vaccinia vector, BSC 1 cells may be infected with the recombinant vector and grown on microscope slides under conditions which allow expression. The cells may then be acetone-fixed, and immunofluorescence assays performed using serum which is known to contain anti-HCV antibodies to a polypeptide(s) encoded in the region of the HCV genome from which the HCV segment in the recombinant expression vector was derived.

25

Other systems for expression of eukaryotic or viral genomes include insect cells and vectors suitable for use in these cells. These systems are known in the art, and include, for example, insect expression transfer vectors derived from the baculovirus *Autographa californica* nuclear polyhedrosis virus (AcNPV), which is a helper-  
30 independent, viral expression vector. Expression vectors derived from this system usually

Sticky-ended cleavage fragments may be blunt ended using *E. coli* DNA polymerase I (Klenow fragment) with the appropriate deoxynucleotide triphosphates (dNTPs) present in the mixture. Treatment with S1 nuclease may also be used, resulting in the hydrolysis of any single stranded DNA portions.

Ligations are carried out under standard buffer and temperature conditions using T4 DNA ligase and ATP; sticky end ligations require less ATP and less ligase than blunt end ligations. When vector fragments are used as part of a ligation mixture, the vector fragment is often treated with bacterial alkaline phosphatase (BAP) or calf intestinal alkaline phosphatase to remove the 5'-phosphate, thus preventing religation of the vector. Alternatively, restriction enzyme digestion of unwanted fragments can be used to prevent ligation.

Ligation mixtures are transformed into suitable cloning hosts, such as *E. coli*, and successful transformants selected using the markers incorporated (e.g., antibiotic resistance), and screened for the correct construction.

Synthetic oligonucleotides may be prepared using an automated oligonucleotide synthesizer as described by Warner, *DNA* (1984) 3:401. If desired, the synthetic strands may be labeled with  $^{32}\text{P}$  by treatment with polynucleotide kinase in the presence of  $^{32}\text{P}$ -ATP under standard reaction conditions.

DNA sequences, including those isolated from cDNA libraries, may be modified by known techniques, for example by site directed mutagenesis (see e.g., Zoller, *Nuc Acids Res* (1982) 10:6487). Briefly, the DNA to be modified is packaged into phage as a single stranded sequence, and converted to a double stranded DNA with DNA polymerase, using as a primer a synthetic oligonucleotide complementary to the portion of the DNA to be modified, where the desired modification is included in the primer sequence. The resulting double stranded DNA is transformed into a phage-supporting host

conserved between the invertebrate cells and vertebrate cells. Examples of the signal sequences from vertebrate cells which are effective in invertebrate cells are known in the art, for example, the human interleukin-2 signal (IL2<sub>s</sub>) which signals for secretion from the cell, is recognized and properly removed in insect cells.

5

Transformation may be by any known method for introducing polynucleotides into a host cell, including, for example packaging the polynucleotide in a virus and transducing a host cell with the virus, and by direct uptake of the polynucleotide. The transformation procedure used depends upon the host to be transformed. Bacterial transformation by direct uptake generally employs treatment with calcium or rubidium chloride (Cohen, Proc Nat Acad Sci USA (1972) 69:2110; T. Maniatis et al, "Molecular Cloning; A Laboratory Manual" (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1982). Yeast transformation by direct uptake may be carried out using the method of Hinnen et al, Proc Nat Acad Sci USA (1978) 75:1929. Mammalian transformations by direct uptake may be conducted using the calcium phosphate precipitation method of Graham and Van der Eb, Virology (1978) 52:546, or the various known modifications thereof. Other methods for introducing recombinant polynucleotides into cells, particularly into mammalian cells, include dextran-mediated transfection, calcium phosphate mediated transfection, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the polynucleotides into nuclei.

Vector construction employs techniques which are known in the art. Site-specific DNA cleavage is performed by treating with suitable restriction enzymes under conditions which generally are specified by the manufacturer of these commercially available enzymes. In general, about 1 mg of plasmid or DNA sequence is cleaved by 1 unit of enzyme in about 20 mL buffer solution by incubation for 1-2 hr at 37°C. After incubation with the restriction enzyme, protein is removed by phenol/chloroform extraction and the DNA recovered by precipitation with ethanol. The cleaved fragments may be separated using polyacrylamide or agarose gel electrophoresis techniques, according to the general procedures described in Meth Enzymol (1980) 65:499-560.

30

following chloramphenicol amplification (Clewell, J Bacteriol (1972) 110:667). The DNA is isolated and analyzed, usually by restriction enzyme analysis and/or sequencing.

Sequencing may be performed by the dideoxy method of Sanger et al, Proc Nat Acad Sci USA (1977) 74:5463, as further described by Messing et al, Nuc Acids Res (1981) 9:309,

5 or by the method of Maxam et al, Meth Enzymol (1980) 65:499. Problems with band compression, which are sometimes observed in GC-rich regions, were overcome by use of T-deazoguanosine according to Barr et al, Biotechniques (1986) 4:428.

The enzyme-linked immunosorbent assay (ELISA) can be used to measure  
10 either antigen or antibody concentrations. This method depends upon conjugation of an enzyme to either an antigen or an antibody, and uses the bound enzyme activity as a quantitative label. To measure antibody, the known antigen is fixed to a solid phase (e.g., a microtiter dish, plastic cup, dipstick, plastic bead, or the like), incubated with test serum dilutions, washed, incubated with anti-immunoglobulin labeled with an enzyme, and  
15 washed again. Enzymes suitable for labeling are known in the art, and include, for example, horseradish peroxidase (HRP). Enzyme activity bound to the solid phase is usually measured by adding a specific substrate, and determining product formation or substrate utilization colorimetrically. The enzyme activity bound is a direct function of the amount of antibody bound.

20

To measure antigen, a known specific antibody is fixed to the solid phase, the test material containing antigen is added, after an incubation the solid phase is washed, and a second enzyme-labeled antibody is added. After washing, substrate is added, and enzyme activity is measured colorimetrically, and related to antigen concentration.

25

The NS3 proteins of the three genera of *Flaviviridae*: flavivirus, pestivirus and HCV, have conserved sequence motifs of serine type proteinase and of nucleoside triphosphatase (NTPase)/RNA helicase. See Figure 2. The NTPase/RNA helicase carboxy two-thirds of the NS3 protein fragment belongs to the DEAD box family. The DEAD box  
30 protein family has eight highly conserved amino acid motifs, one of which is the DEAD

bacterium. Cultures of the transformed bacteria which contain copies of each strand of the phage are plated in agar to obtain plaques. Theoretically, 50% of the new plaques contain phage having the mutated sequence, and the remaining 50% have the original sequence. Replicates of the plaques are hybridized to labeled synthetic probe at temperatures and  
5 conditions which permit hybridization with the correct strand, but not with the unmodified sequence. The sequences which have been identified by hybridization are recovered and cloned.

DNA libraries may be probed using the procedure of Grunstein and  
10 Hogness Proc Nat Acad Sci USA (1975) 73:3961. Briefly, in this procedure the DNA to be probed is immobilized on nitrocellulose filters, denatured, and prehybridized with a buffer containing 0-50% formamide, 0.75 M NaCl, 75 mM Na citrate, 0.02% (wt/v) each of bovine serum albumin, polyvinylpyrrolidone, and Ficoll®, 50 mM NaH<sub>2</sub>PO<sub>4</sub> (pH 6.5),  
15 0.1% SDS, and 100 mg/mL carrier denatured DNA. The percentage of formamide in the buffer, as well as the time and temperature conditions of the prehybridization and subsequent hybridization steps depend on the stringency required. Oligomeric probes which require lower stringency conditions are generally used with low percentages of formamide, lower temperatures, and longer hybridization times. Probes containing more than 30 or 40 nucleotides, such as those derived from cDNA or genomic sequences  
20 generally employ higher temperatures, e.g., about 40-42°C, and a high percentage formamide, e.g., 50%. Following prehybridization, 5'-<sup>32</sup>P-labeled oligonucleotide probe is added to the buffer, and the filters are incubated in this mixture under hybridization conditions. After washing, the treated filters are subjected to autoradiography to show the location of the hybridized probe; DNA in corresponding locations on the original agar plates is used  
25 as the source of the desired DNA.

For routine vector constructions, ligation mixtures are transformed into *E. coli* strain HB101 or other suitable hosts, and successful transformants selected by antibiotic resistance or other markers. Plasmids from the transformants are then prepared  
30 according to the method of Clewell et al, Proc Nat Acad Sci USA (1969) 62:1159, usually

was inhibited by a NS3-specific monoclonal antibody, but, an unrelated antibody did not affect the activity. From these results, it was determined that the helicase activity was derived not from *E. coli* contaminants, but from the HCV NS3 protein fragments.

- 5                   Most of the investigated RNA helicases bound to single strand region and then unwound double strand RNA by moving unidirectionally or bidirectionally. The substrate with the single strand region on both 3' and 5' ends was used. Suzich *et al.*, J. Virol., 67:6152-6158 (1993) showed that the two thirds of the C'-terminal of HCV NS3 could hydrolyze all NTPs and dNTPs. This NTPase activity was observed with the HCV  
10 NS3 protein fragments disclosed herein. (data not shown) The results showing that the truncated NS3 protein fragments described herein having biochemical helicase activity in spite of deleted N<sup>4</sup>-terminal proteinase domain suggest that the proteinase and NTPase domains may act independently.
- 15                   The HCV NS3 protein fragments showing helicase activity of the present invention are advantageous because they are soluble in purification and assay buffers, while the entire NS3 protein generally is not. The solubility of the helicase fragments was determined by first constructing several clones from various vectors and fusion proteins. For example, a pGEX-2T vector containing a glutathione-s-transferase (GST) fusion  
20 protein was used to clone the HCV NS3 protein i.e., from 1027 to 1657 a.a. of HCV-1. The resulting fusion protein of GST and HCV NS3 protein was insoluble, i.e., the only portion of the fusion protein that was isolated was that from the insoluble portion of the bacterial extract. That fusion protein was solubilized by denaturing with 6 M urea. When the denatured fusion protein was refolded by serial dialysis against a concentration step  
25 gradient, only a small fraction of the renatured fusion protein was correctly refolded and no enzymatic activity was observed in the renatured protein. When an HCV NS3 protein was fused with a maltose binding protein using a pMAL vector, the fusion protein was soluble. The molecular weight of the fusion protein, however, was relatively large (M.W. 110 kDa) because the maltose binding protein itself is about 40 kDa. Thus, such a fusion protein is  
30 undesirable to use. In addition, it is difficult to separate the maltose binding protein domain

region where it is also known as an ATPase motif. The DEAD protein family consists of three subfamilies: DEAD proteins, DEAH proteins and DEXH proteins. Figure 3 shows the conserved sequence motifs of DEXH protein family and the corresponding motifs of HCV NS3. The HCV NS3 protein has sequence motif of DECH which results in its  
5 classification in the DEXH protein subfamily.

The HCV NS3 protein fragments disclosed herein have similar characteristics with other known RNA helicases, i.e., they show RNA helicase activity only in the presence of divalent cations ( $Mn^{2+}$  or  $Mg^{2+}$ ) and ATP. At a lower level of ATP,  
10 (approximately 1 mM) an increasing amount of either cation inhibits the enzymatic activity of the NS3 fragment. When the ATP concentration is high, (approximately 5 mM), helicase activity remains at a high level even when  $Mg^{2+}$  or  $Mn^{2+}$  cations are present at high concentrations. RNA helicase A purified from HeLa cells, needs only  $Mg^{2+}$  for its cofactor, and  $Mn^{2+}$  does not substituted for  $Mg^{2+}$ . See Lee et al., *J. Biol.* 267:4398-4407  
15 (1992), incorporated herein by reference. Pestivirus NS3 and Vaccinia virus RNA helicase have shown to use both cations. Likewise, HCV NS3 protein helicase fragments disclosed herein can utilize both metal ions.

The helicase activity of the HCV NS3 protein helicase fragments is likely  
20 pH specific. The experiments in the examples were carried out at pH 6.5. When the pH was increased to 7.6, however, HCV NS3 protein helicase fragments showed not more than 10% strand separation, keeping all other components constant. (data not shown) These characteristics of HCV NS3 protein helicase fragments imply that it has a similar nature to pestivirus NS3 RNA helicase, which is known to pH sensitive.

25  
RNA helicase activity was confirmed not to be derived from *E. coli* contaminants in two ways. First, a pET21b plasmid without a HCV NS3 protein fragment insert was used as a negative control. The enzymatic activity of the same eluted fraction from the negative control cell culture was tested and there was no detectable level of  
30 NTPase or RNA helicase activity. Second, the NS3 protein fragment's helicase activity

polyacrylamide gel by studying the unwinding activity of the helicase fragment. The screening assay can also be automated in a microtiter dish (96-well plate) format. In the latter assay, the double-stranded template RNA is labeled with biotin at the 5'-end of one strand and with  $^{32}\text{P}$  at the 5'-end of the other strand. This labeled template can be attached  
5 to the bottom of the well that is coated with streptavidin. The helicase activity from the added fragments can be measured by counting radioactivity from the displaced  $^{32}\text{P}$ -labeled RNA strand that is now present in the well supernatant. Potential helicase inhibitors present in the combinatorial library can be found by detecting specific inhibition of the strand displacement reaction by helicase fragments.

10

### C. Examples

The examples presented below are provided as a further guide to the practitioner of ordinary skill in the art, and are not to be construed as limiting the invention in any way.

15

#### Example 1

Expression and purification of HCV NS3 protein. For expressing the carboxy two-thirds of HCV NS3 protein, the polymerase chain reaction (PCR) was used to amplify a 1.4 Kb  
20 DNA fragment encompassing amino acids 1193 to 1657 from HCV-1 cDNA. The sense primer used was JCK-1 5'-GGGGATCCGGTGGACTTTATCCCT-3' (SEQ ID NO: 4), and the antisense primer JCK-7 5'-GGAAGCTTGCTGACGACCTCG-3' (SEQ ID NO: 5). The PCR produced was digested with BamHI and HindIII inserted into BamHI and HindIII sites of pET21b (purchased from Novagen, WI).<sup>1</sup> The recombinant plasmid was  
25 designated as pET21b-NS3HCV and transformed to *E. coli* BL21 (DE3), and the inserted region was verified by sequencing. pET21b-NS3HCV consisted of 466 amino acid

---

<sup>1</sup> As a negative control, a pET21b plasmid without the insert was transformed to *E. coli* BL21 (DE3) and induced with 1 mM IPTG. The negative control cell culture was processed with the same purification step as pET21b-NS3HCV. The negative control showed no enzymatic activity. See Figure 6, lane 1.



out from the fusion protein containing it and the HCV NS3 protein. In addition, a pET21b vector was utilized to express the domain of HCV NS3 protein, amino acids 1027 to 1657. The expression level of the protein was very low and only a small quantity of the protein was isolated.

5                    Thus, the HCV NS3 protein fragments of the present invention in, e.g., a pET vector system, provides the following advantages:

- 1)            a better T7 promoter system when compared to the promoters of pMAL or pGEX vector;
- 10            2)            an increase in solubility of the expressed NS3 protein fragment having helicase activity;
- 3)            an elimination of the necessity to remove the non-HCV NS3 protein fragment from the fusion protein; and
- 4)            a convenient purification step by using nickel column
- 15            chromatography.

Further, a soluble NS3 protein fragment having helicase activity has several advantages to the insoluble full length protein. First, it is not necessary for the soluble protein fragments to denature and refold for use in purification and enzyme assays. An insoluble protein or  
20            fragment needs to be denatured by urea or Guanidium-HCl for purification and then must be dialyzed against a concentration step gradient for removing the urea or Guanidium-HCl before refolding and recovery of the enzymatic activity of the protein fragment. Second, the yield of soluble NS3 protein fragments from expression systems is higher than that of insoluble NS3 protein fragments. During the denaturation-refolding process, an insoluble  
25            protein fragment is lost in a large portion of the cell extract. Third, the enzymatic activity of the insoluble NS3 proteins cannot be observed after refolding.

Soluble helicase fragments of a HCV NS3 protein can be used to screen for specific helicase inhibitors from a combinatorial library. The screening assay can be  
30            performed based on the mobility shift of the double stranded template RNA in a

was labeled with [ $\alpha$ - $^{32}$ P]-CTP, and the specific activity of labeled substrated was 1 - 1.5 x 10<sup>5</sup> cpm/pmol ds RNA substrate.<sup>3</sup> Duplex RNA was electrophoresed on 6% native polyacrylamide gel (30:0.8), and the location of the ds RNA was identified by autoradiography. To recover the RNA substrate, a sliced gel fragment was ground in 400  $\mu$ l of elution buffer (0.5 M ammonium acetate, 0.1% SDS, 10 mM EDTA) and shaken vigorously at 4°C for 2 hrs. The supernatants were extracted with chloroform and precipitated with ethanol, and the RNA pellet was dissolved in D.W.

### Example 3

10

RNA helicase assay. An RNA helicase assay was performed in 20  $\mu$ l of reaction mixture: 1 pmol NS3 HCV protein fragment, 0.5 pmol ds RNA substrate, 25 mM MOPS-KOH (pH 6.5), 5mM ATP, 3 mM MnCl<sub>2</sub>, 2 mM DTT, 100  $\mu$ g/ml BSA, and 2.5 U RNasin (Promega). The reaction mixture was incubated at 37°C for 30 min. The reaction was stopped by adding 5  $\mu$ l of 5 x termination buffer [0.1 M Tris-Cl (pH 7.4), 20 mM EDTA, 0.5% SDS, 0.1% NP-40, 0.1% bromophenol blue, 0.1% xylene cyanol, and 50% glycerol]. Each aliquot was loaded on 6% native polyacrylamide gel (30:0.8) and electrophoresed at 80 V for 3 hr. The ds RNA substrate and unwound RNA strand were visualized by autoradiography. The effect of ATP and divalent metal ion on the NS3 protein fragment's helicase activity was investigated by carrying out the same reactions with 1, 2, 3, 4, and 6 mM Mn<sup>2+</sup> or Mg<sup>2+</sup> in the presence of 1 mM or 5 mM ATP. Strand separation efficiencies were calculated by counting the radioactivities of the bands with PhosphorImager (Molecular Dynamics, Sunnyvale, CA). See Fig. 7 for the activity change of the HCV NS3 protein fragments at various concentrations of ATP and the divalent cations. The HCV NS3 RNA helicase fragments required divalent ions such as Mg<sup>2+</sup> and Mn<sup>2+</sup> (See Fig. 6, lane 2 to 5). Strand displacement was observed only when Mg<sup>2+</sup> or Mn<sup>2+</sup> ions were present (See Fig. 6, lanes 2 and 4). When either these divalent cations or ATP was deleted, ds RNA was not unwound (See Fig. 6, lanes 3, 5, and 7). Monovalent potassium ion did

<sup>3</sup> Strand displacement were observed by band shift of the radiolabeled long strand.

residues from the carboxy terminus of HCV NS3 and contained His-Tag (6 histidines) and 19 additional residues from the pET expression vector at C-terminal end for easier purification. About 54 kDa (481 amino acid residues) of HCV NS3 His-tag fusion protein was induced by 1 mM IPTG from *E. coli* BL21 (DE3) harboring the recombinant plasmid to exponentially grow cells in LB medium with 10 mg/ml of ampicillin. (See Fig. 5, lanes 1 and 2).<sup>2</sup> From 200 ml of the culture, 400 mg of protein of approximately 95% purity was obtained. After 3 hrs of culturing at 37°C, the cells were harvested and disrupted. Soluble parts of cell extract were loaded onto a metal-binding column. Resin-bound protein was eluted with 1 M imidazole, 0.5 M NaCl, 20 mM Tris-Cl pH 7.9. Eluted fractions were subjected to SDS-PAGE, and protein-containing fractions were pooled and dialyzed against 50 mM Tris-Cl pH 7.9 for 4 hrs. The NTPase assay on polyethyleneimine cellulose TLC (J.T.Baker) was performed as previously described in *Suzich et al.*, to confirm that final purified protein had active conformation. The purified protein showed an NTPase activity (data not shown).

15

### Example 2

Preparation of substrate for RNA Helicase. Fig. 4 shows the structure of the double strand RNA used as a substrate of an RNA helicase. The long strand was prepared by *in vitro* transcription of pGEM1 that had been cleaved with PvuII, and the short strand was transcribed from the BamHI digested pSP65. Both strands were transcribed with SP6 RNA polymerase (New England Biolabs) according to the manufacturer's manual. After the transcription reaction, each aliquot was treated with RNase-free DNase (Promega) and extracted with phenol:chloroform, and precipitated with ethanol. Each RNA strand was resuspended with 25 ml of hybridization buffer (20 mM HEPES-KOH pH 7.6, 0.5 M NaCl, 1 mM EDTA, 0.1% SDS), and mixed. The mixture was heated to 100°C for 5 min. and incubated at 65°C for 30 min. and incubated at 25°C overnight. The long strand RNA

<sup>2</sup> (One or more protein bands about 50 kDa appeared by IPTG induction, but only the 54 kDa NS3-His fusion protein was purified from the metal binding affinity column. (See Fig. 5, lane 3))

NS3 helicase domain, i.e., amino acid 1193 to amino acid 1608 of the HCV NS3 domain, ATCC deposit no. 97309; No. 5, an HCV NS3 fragment having 97 amino acids deleted from the C-terminus of the HCV NS3 helicase domain, i.e., amino acid 1193 to amino acid 1651 of the HCV NS3 domain, ATCC deposit no. 97310; No. 6, an HCV NS3 fragment  
5 having 135 amino acids deleted from the C-terminus of the HCV NS3 helicase domain, i.e., amino acid 1209 to amino acid 1657 of the HCV NS3 domain, ATCC deposit no. 97311; No. 7, an HCV NS3 fragment having 16 amino acids deleted from the N-terminus of the HCV NS3 helicase domain, i.e., from amino acid 1209 to amino acid 1657 of the HCV NS3 domain, ATCC deposit no. 97312; and No. 8, an HCV NS3 fragment having  
10 32 amino acids deleted from the N-terminus of the HCV NS3 helicase domain, i.e., from amino acid 1225 to amino acid 1657 of the HCV NS3 domain, ATCC deposit no. 97313

As shown in Fig 9, truncated mutants, numbers 5, 6, and 8 mutants did not demonstrate RNA helicase activity. Mutant 7, however, did demonstrate NTPase  
15 activity even though its activity was about half of No. 1 (full length) protein. Fig. 8 shows the RNA helicase assay of truncated mutants. The upper band indicates dsRNA and the lower band ssRNA labelled with <sup>32</sup>P. Boiled RNA indicates denatured dsRNA after boiling for 5 min, and was therefore a control for ssRNA. As shown in both Figs 8 and 9, truncated fragments numbers 5, 6, and 8 lost RNA helicase activity.

20

#### Example 5

Determining solubility of the HCV NS3 fragments The solubility of the expressed protein from pET21b-HCVNS3 vector was determined by the following method: ITPG-  
25 induced cells were harvested at 6000 G for 5 mins. The cells were then resuspended with 1X binding buffer (5 mM imidazole, 500 mM NaCl, 20 mM Tris-Cl pH 7.9). The resuspended cells were then frozen in a dry ice-ethanol bath and thawed on ice and sonicated for 2 min. Cell extracts were centrifuged at 27000 G for 30 min. The soluble part of the cell extract, the supernatant and the insoluble part of the cell extract, the  
30 pellet, were subjected on SDS-PAGE. When a western blot was carried out for the SDS-

not activate the HCV NS3 protein fragment's helicase activity at these conditions (See Figure 6, lane 6). At 1 mM ATP, the helicase activity was lower than at 5mM (See Figure 6, lane 8). Enzymatic activity of NS3 was inhibited by monoclonal antibodies of HCV NS3 protein fragments (See Figure 6, lane 9), and was not blocked by a non-specific  
5 antibody at two different concentrations (See Figure 6, lanes 10 and 11).

As mentioned above, RNA helicase activity of the HCV NS3 protein fragments was dependent on divalent cations and ATP. At low concentration of ATP (1 mM), helicase activity of NS3 was highest at a low concentration of either of the divalent  
10 cations, and, the helicase activity decreased when the concentration of the cations was increased. At high concentration of ATP (5 mM), most of the substrates were unwound at all of the tested cation concentrations. At 3 mM or 4 mM of cation concentration, either  $Mn^{2+}$  or  $Mg^{2+}$ , the helicase activity was the highest. Thus, the helicase activity appears more sensitive to the divalent cation concentration in lower concentrations of ATP. In  
15 addition, the HCV NS3 protein fragments showed a slight bias for  $Mg^{2+}$ .

#### Example 4

Testing of Truncated HCV NS3 Fragments for Helicase Activity HCV NS3 fragments of  
20 varying sizes were expressed and purified as described above. The fragments were then tested for helicase activity as described above, and for NTPase activity as is known in the art. Fig. 9 depicts the fragments tested and whether the fragments showed helicase/NTPase activity. The following fragments were tested: No. 1, a full length helicase fragment, i.e., from amino acid 1193 to amino acid 1657 of the HCV NS3 domain, ATCC deposit  
25 no.97306; No. 2, an HCV NS3 fragment having 10 amino acids deleted from the C-terminus of the HCV NS3 helicase domain, i.e., from amino acid 1193 to amino acid 1648 of the HCV NS3 domain, ATCC deposit no. 97307; No. 3, an HCV NS3 fragment having 30 amino acids deleted from the C-terminus of the HCV NS3 helicase domain, i.e., amino acid 1193 to amino acid 1638 of the HCV NS3 domain, ATCC deposit no. 97308; No. 4,  
30 an HCV NS3 fragment having 50 amino acids deleted from the C-terminus of the HCV

**WHAT IS CLAIMED IS:**

1. A composition comprising an HCV NS3 helicase fragment having substantially the same amino acid sequence as shown in SEQ ID NO:2.
2. The composition of claim 1 wherein said fragment is produced by chemical synthesis or recombinant DNA expression.
3. A composition according to either of claims 1 or 2 wherein said fragment has its carboxy terminus at amino acid 1561 to 1657.
4. A composition according to any of claims 1-3, wherein said fragment has its amino acid terminus at amino acid 1193 to amino acid 1223.
5. A fusion protein comprising a suitable fusion partner fused to a HCV NS3 helicase fragment having substantially the amino acid sequence as shown in SEQ ID NO:2.
6. The fusion protein of claim 5 wherein said fusion partner comprises human superoxide dismutase.
7. The fusion protein according to either of claims 5 or 6 wherein said fragment is produced by chemical synthesis or recombinant DNA expression.
8. The fusion protein according to any of claims 5-7 wherein said helicase fragment has its carboxy terminus at amino acid 1561 to amino acid 1657.

PAGE using a monoclonal antibody against the HCV NS3 protein fragment, the expressed protein was observed only in the soluble part of the cell extract.

The above materials deposited with the ATCC under the accession  
5 numbers indicated, will be maintained under the terms of the Budapest Treaty on the  
International Recognition of the Deposit of Microorganisms for purposes of Patent  
Procedure. These deposits are provided as a convenience to those of skill in the art, and  
are not an admission that a deposit is required under 35 U.S.C. §112. The polynucleotide  
sequences contained in the deposited materials, as well as the amino acid sequence of the  
10 polypeptides encoded thereby, are incorporated herein by reference and are controlling in  
the event of any conflict with the sequences described herein. A license may be required  
to make, use or sell the deposited materials, and no such license is granted hereby.

contacting said fragment with a compound capable of inhibiting RNA helicase activity; and  
measuring inhibition of the activity of said hepatitis C virus helicase.

17. The method of claim 16 wherein said fragment has its carboxy terminus at amino acid 1561 to amino acid 1657.

18. The method according to either of claims 16 or 17 wherein said fragment has its amino terminus at amino acid 1193 to amino acid 1223.

19. An expression vector for producing HCV NS3 active helicase fragments in a host cell, which vector comprises:

a polynucleotide encoding a HCV NS3 helicase fragment  
having having substantially the same amino acid sequence shown in SEQ ID NO:2;  
transcriptional and translational regulatory sequences  
functional in said host cell operably linked to said fragment-encoding polynucleotide;  
and

a selectable marker.

20. The vector of claim 19 further comprising a sequence encoding a fusion partner, linked to said fragment-encoding polynucleotide to form a fusion protein upon expression.

21. The vector of claim 20 wherein said fusion partner is selected from the group consisting of hSOD, yeast alpha-factor, IL-2S, ubiquitin, beta-galactoside, beta-lactamase, horseradish peroxidase, glucose oxidase, and urease.

22. The vector according to any of claims 19-21, wherein said fragment has its carboxy terminus at amino acid 1561 to amino acid 1657.



9. The fusion protein according to any of claims 5-8 wherein said helicase fragment has its amino terminus at amino acid 1193 to amino acid 1223.

10. The fusion protein according to claim 5 wherein said fusion partner is ubiquitin.

11. A composition comprising a polynucleotide which encodes a HCV NS3 helicase fragment having substantially the amino acid sequence of SEQ ID NO:2.

12. A composition comprising a polynucleotide which encodes a fusion protein comprising a HCV NS3 helicase fragment having substantially the amino acid sequence of SEQ ID NO:2 and a fusion partner.

13. The composition according to claim 12 wherein said fusion partner is selected from the group consisting of hSOD, yeast alpha-factor, IL-2S, ubiquitin, beta-galactoside, beta-lactamase, horseradish peroxidase, glucose oxidase, and urease.

14. The composition according to either of claims 11 or 12, wherein said helicase fragment has its carboxy terminus at amino acid 1561 to amino acid 1657.

15. The composition according to any of claims 11-14, wherein said helicase fragment has its amino terminus at amino acid 1193 to amino acid 1223.

16. A method for assaying compounds for activity against hepatitis C virus comprising:  
providing an HCV NS3 helicase fragment having substantially the same sequence shown in SEQ ID NO:2;

23. The vector according to any of claims 19-22, wherein said fragment has its amino terminus at amino acid 1193 to amino acid 1223.

APITAYAQQTRGLLGCIITSLTGR  
DKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTIA SPKGPVIQM-1100

S T  
YTNVDQDLVGWPAPQGSRS LTPCTCGSSDLYLVTRHADVIPVRRRGDSRG  
SLLSPRPISYLGSSGGP L LCPAGHAVGIFRAAVCTRGVAKAVDFIPVEN-1200  
LETTMRSPVFTDNSSPPVVPQS FQVAHLHAPT GSGKSTKVPAAYAAQGYK

L  
VLVLNPSVAATLGFGAYMSKAHGIDPNIR TGVRTITTGSPITYSTYGKFL-1300  
ADGGCSGGAYDIIICDECHSTDATSILGIGTVLDQAETAGARLVVLATAT  
PPGSVTVPHPNIEEVALSTTGEIPFYGKAIPLEVIKGGRHLIFCHSKKKC-1400  
DELA AKLVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTGYTGDFDS

Y (S)  
VIDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTQRRGRTGRGKPGIYR-1500  
FVAPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVRLRAYMNT PGLPV  
CQDHLEFWEGVFTGLTHIDAHFLSQTQSGENLPYL VAYQATVCARAQAP-1600  
PPSWDQMWKCLIRLKPTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTCMS  
ADLEVVT

FIGURE 1

2/6

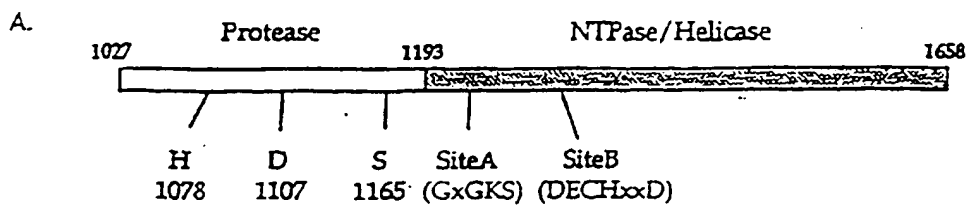


FIGURE 2

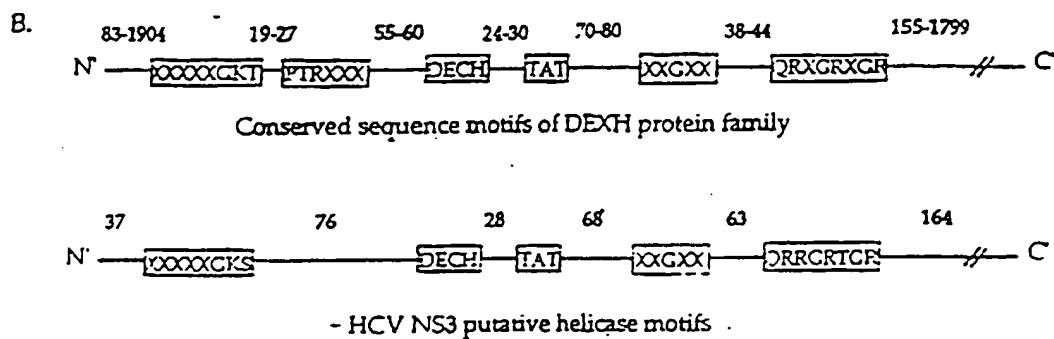


FIGURE 3

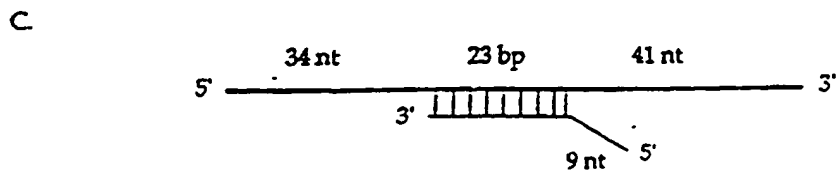


FIGURE 4

3/6

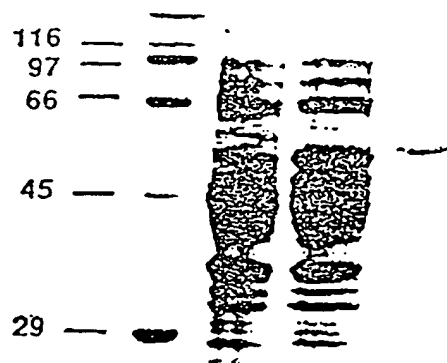


FIGURE 5



FIGURE 6

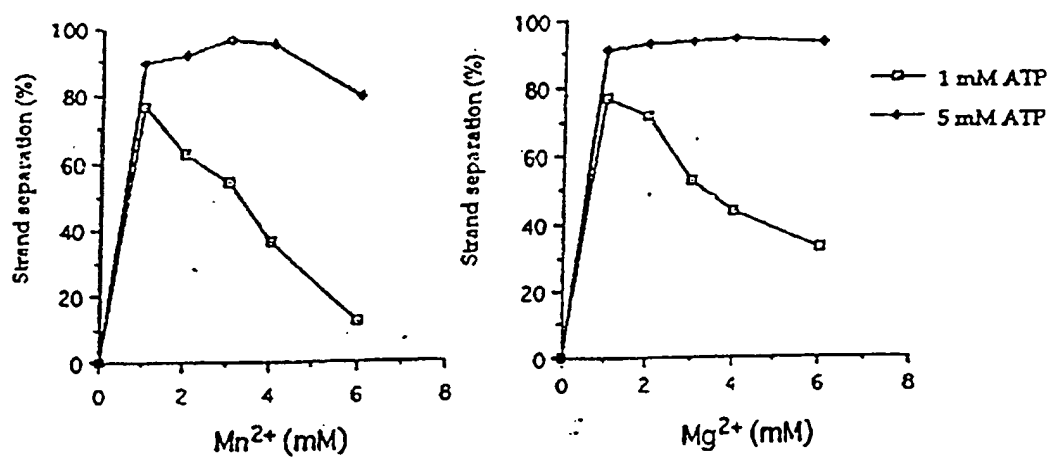
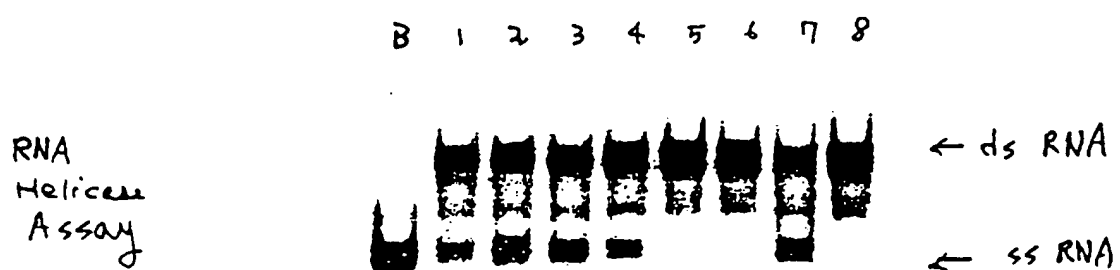


FIGURE 7

5/6



B. boiled RNA. (= ss RNA)

FIGURE 8

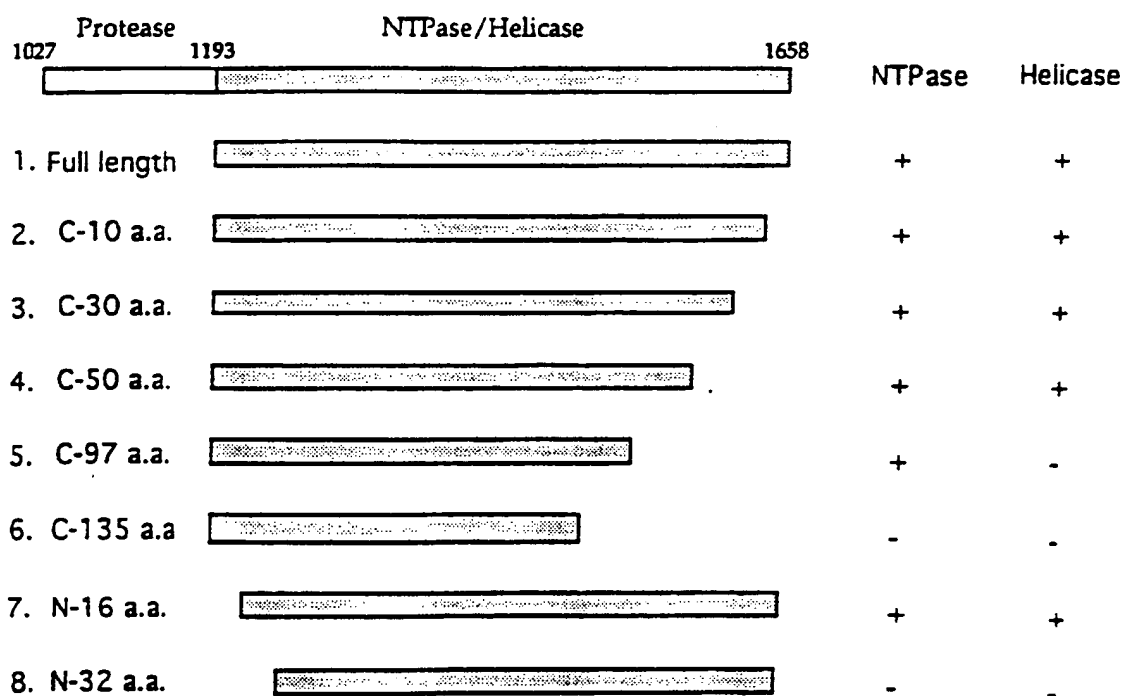


FIGURE 9